

Distributed Detection via Bayesian Updates and Consensus

LIU Qipeng¹, ZHAO Jiuhua², WANG Xiaofan²

1. Institute of Complexity Science, Qingdao University, Qingdao 266071, P. R. China
E-mail: qipengliu@qdu.edu.cn

2. Department of Automation, Shanghai Jiao Tong University, and
Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, P. R. China
E-mail: {jiuhuadandan,xfwang}@sjtu.edu.cn

Abstract: In this paper, we discuss a class of distributed detection algorithms which can be viewed as implementations of Bayes' law in distributed settings. Some of the algorithms are proposed in the literature most recently, and others are first developed in this paper. The common feature of these algorithms is that they all combine (i) certain kinds of consensus protocols with (ii) Bayesian updates. They are different mainly in the aspect of the type of consensus protocol and the order of the two operations. After discussing their similarities and differences, we compare these distributed algorithms by numerical examples. We focus on the rate at which these algorithms detect the underlying true state of an object. We find that (a) The algorithms with consensus via geometric average is more efficient than that via arithmetic average; (b) The order of consensus aggregation and Bayesian update does not apparently influence the performance of the algorithms; (c) The existence of communication delay dramatically slows down the rate of convergence; (d) More communication between agents with different signal structures improves the rate of convergence.

Key Words: Networked Systems, Distributed Detection, Consensus, Bayes' Law

1 Introduction

Recent years have witnessed a considerable amount of work on analysis of networked systems ranging from social and economic networks to robot and sensor networks [1–4]. An amazing phenomenon arising in networked systems is that, by communicating and cooperating among individuals, the whole group could complete very complicated tasks way beyond the ability of any single agent.

A common task in networked systems is that all agents are supposed to collectively find out an underlying true state of an object using relatively local information such as private observations and neighbors' information. For instance, voters attempt to find out the ability of some political candidates; costumers learn the quality of a new product; a network of sensors detect the mean temperature of a wide area. According to specific contexts, this task might have different names, such as social learning, distributed detection, distributed estimation, and distributed hypothesis testing [5–14]. To be consistent, we call this task *distributed detection* throughout this paper. The aim of the whole group is to detect the underlying true state of an object. To complete the task, a variety of distributed algorithms are designed in the literature to effectively aggregate information scattered all over the network.

In this paper, we focus on a class of distributed detection algorithms which involve the implementation of Bayes' law in a distributed setting. It is well known that the standard Bayes' law is very useful in detection, estimation, hypothesis testing, and other similar applications. By continuously observing new data or other useful information, an individual could eventually learn the true state. In a networked setting, it is a common

phenomenon that any single agent could not learn the true state by itself. However, communicating with others in the network might bring the agents more useful information, and under certain conditions, all agents might eventually learn the true state collectively.

Our work in this paper is directly motivated by [10–14] where different but quite similar distributed detection algorithms are developed. All of these algorithms combine certain kinds of consensus protocols with Bayesian updates. They are different mainly in the aspect of the type of consensus protocol and the order of the two operations. In [10], agents first logarithmically aggregate their neighbors' beliefs to form a new prior belief, and then update their own beliefs using Bayes' law. In [11], in contrast to that in [10], an distributed detection algorithm with local Bayesian update first is proposed, i.e., the two operations change their order. In stead of logarithmically aggregating neighbors' posterior beliefs like that in [11], an algorithm with linearly aggregation is proposed in [12]. In [13], each agent first compute its Bayesian posterior belief based on its private observation, and then linearly combine it with neighbors' prior beliefs which can be interpreted as communication delay in the dynamics. The distributed detection rule in [14], also involving delayed communication, logarithmically combines the Bayesian posterior based on private observation and neighbors' prior beliefs, which is can be viewed as a logarithmic analog of the rule in [13].

In this paper, we first provide a systemic discussion about the class of distributed detection algorithms which combine certain kinds of consensus protocols with Bayesian updates, and show their essential similarities and differences. Some of the algorithms are proposed in the literature most recently, and others are first proposed in this paper. Then, we provide numerical examples to compare their performance in distributed detection problems. Some qualitative results are given which

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61374176 and 61473189, and the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (No. 61221003).

might provide us insight into the design of more efficient distributed detection algorithms.

This paper is organized as follows. In Sec. 2 we discuss and compare the distributed detection algorithms. In Sec. 3 we provide numerical examples to analyze the factors which influence the efficient of distributed detection algorithms. Concluding remarks are given in Sec. 4.

2 Distributed Detection Algorithms Based on Consensus and Bayesian Updates

2.1 Preliminaries

Consider a social network as a directed graph $G = (V, E)$, where $V = \{1, 2, \dots, n\}$ is the node set and $E \subset V \times V$ is the edge set. Each node in V represents an agent, and the edge from i to j , denoted by the order pair $(i, j) \in E$, captures the fact that agent i is a neighbor of agent j , and j can receive some information from i . The set of neighbors of agent i is denoted by $N_i = \{j \in V : (j, i) \in E\}$. Moreover, weight $a_{ij} \in [0, 1]$ is assigned to any ordered pair of agents such that $a_{ij} > 0$ if and only if $j \in N_i$. The weight $a_{ii} \geq 0$ is the self-weight of agent i , and we posit that $\sum_{i=1}^n a_{ij} = 1$.

Let θ denote a state of the object we are concerned with, and all the possible states compose a state set $\Theta = [\theta_1, \theta_2, \dots, \theta_m]$, in which the true state is denoted by θ^* . From the point of view of agent i at time t , the probability of state θ being true is denoted by $\mu_{i,t}(\theta)$, which is called the *belief* of agent i on θ . Thus, agent i 's belief $\mu_{i,t} = [\mu_{i,t}(\theta_1), \mu_{i,t}(\theta_2), \dots, \mu_{i,t}(\theta_m)] \in \mathbb{P}(\Theta)$ is a probability distribution over Θ , where $\mathbb{P}(\Theta)$ is the set of all possible probability distribution over Θ .

Conditional on the underlying true state, at each time period $t > 0$, a signal vector $s_t = (s_{1,t}, s_{2,t}, \dots, s_{n,t}) \in S$ is generated according to the likelihood function $\ell(s_t|\theta^*)$, where $s_{i,t}$ is the signal observed by agent i and S is the signal space. For each observed signal s and each possible state θ , agent i holds a corresponding private signal structure $\ell_i(s|\theta) > 0$, representing the probability that it believes signal s appears if the true state is θ . We assume that the private signal structure of agent i about the true state θ^* is the i -th marginal of $\ell(\cdot|\theta^*)$, which means the agent has a perfect prior information about the true state. If there exists a state $\bar{\theta} \neq \theta^*$ satisfying that $\ell_i(s|\bar{\theta}) = \ell_i(s|\theta^*)$ for all signal s , we call $\bar{\theta}$ observationally equivalent to the true state. That is to say, state $\bar{\theta}$ and the underlying true state θ^* arouse exactly the same signals according to the same probability in agent i 's eyes, and thus, he cannot tell these two states apart only by observing the signals. All the states that observationally equivalent to θ^* from the point of view of agent i compose a set $\bar{\Theta}_i = \{\theta \in \Theta : \ell_i(s|\theta) = \ell_i(s|\theta^*), \forall s \in S\}$. If $\cap_{i=1}^n \bar{\Theta}_i = \{\theta^*\}$, we say the true state θ^* is globally identifiable.

In the next, we will describe six distributed detection algorithms used to detect the underlying true state, which all can be viewed as combinations of consensus protocols and Bayesian updates.

2.2 Interpreting the Bayesian Posterior as the Solution of an Optimization Problem

The standard Bayesian posterior obtain by agent i based on its observation $s_{i,t+1}$ is as follows:

$$\mu_{i,t+1}(\theta) = \frac{\mu_{i,t}(\theta)\ell_i(s_{i,t+1}|\theta)}{\sum_{\theta_k \in \Theta} \mu_{i,t}(\theta_k)\ell_i(s_{i,t+1}|\theta_k)}, \quad \theta \in \Theta. \quad (1)$$

As point out in [10, 15, 16], the posterior belief can be interpreted as the solution of the following optimization problem

$$\mu_{i,t+1} = \arg \min_{\pi \in \mathbb{P}(\Theta)} \left\{ D_{KL}(\pi \| \mu_{i,t}) - \sum_{\theta_k \in \Theta} \pi(\theta_k) \log(\ell_i(s_{i,t+1}|\theta_k)) \right\} \quad (2)$$

where $D_{KL}(\pi \| \mu_{i,t})$ is the Kullback-Leibler divergence (the KL-divergence for short) between probability distributions π and $\mu_{i,t}$ with the following definition

$$D_{KL}(\pi \| \mu_{i,t}) = \sum_{\theta_k \in \Theta} \pi(\theta_k) \log \frac{\pi(\theta_k)}{\mu_{i,t}(\theta_k)}.$$

Note that the first term on the right hand side of (2) measures the difference between the distributions π and $\mu_{i,t}$, and the second term is the maximum likelihood estimation given the observation $s_{i,t+1}$. Thus, the posterior distribution can be viewed as a tradeoff between the prior belief and the observation.

In a network setting, by introducing neighbors' prior distribution into the optimization problem (2), we obtain the following new optimization problem

$$\mu_{i,t+1} = \arg \min_{\pi \in \mathbb{P}(\Theta)} \left\{ \sum_{j=1}^n a_{ij} D_{KL}(\pi \| \mu_{j,t}) - \sum_{\theta_k \in \Theta} \pi(\theta_k) \log(\ell_i(s_{i,t+1}|\theta_k)) \right\}. \quad (3)$$

Note that

$$\begin{aligned} & \sum_{j=1}^n a_{ij} D_{KL}(\pi \| \mu_{j,t}) \\ &= \sum_{j=1}^n a_{ij} \sum_{\theta_k \in \Theta} \pi(\theta_k) \log \frac{\pi(\theta_k)}{\mu_{j,t}(\theta_k)} \\ &= \sum_{\theta_k \in \Theta} \pi(\theta_k) \sum_{j=1}^n a_{ij} \log \frac{\pi(\theta_k)}{\mu_{j,t}(\theta_k)} \\ &= \sum_{\theta_k \in \Theta} \pi(\theta_k) \log \prod_{j=1}^n \left(\frac{\pi(\theta_k)}{\mu_{j,t}(\theta_k)} \right)^{a_{ij}} \\ &= \sum_{\theta_k \in \Theta} \pi(\theta_k) \log \frac{\pi(\theta_k)}{\prod_{j=1}^n \mu_{j,t}^{a_{ij}}(\theta_k)} \\ &= D_{KL} \left(\pi \| \prod_{j=1}^n \mu_{j,t}^{a_{ij}} \right) \end{aligned}$$

which is the KL-divergence between π and the geometric mean of the prior beliefs of agent i and its neighbors.

By the above derivation, the optimization problem (3) can be rewritten as

$$\mu_{i,t+1} = \arg \min_{\pi \in \mathbb{P}(\Theta)} \left\{ D_{KL} \left(\pi \parallel \prod_{j=1}^n \mu_{j,t}^{a_{ij}} \right) - \sum_{\theta_k \in \Theta} \pi(\theta_k) \log(\ell_i(s_{i,k+1} | \theta_k)) \right\}. \quad (4)$$

The solution of (3) (also (4)), is the Bayesian posterior belief corresponding to the prior $\prod_{j=1}^n \mu_{j,t}^{a_{ij}}$, which has the following form

$$\mu_{i,t+1}(\theta) = \frac{\prod_{j=1}^n \mu_{j,t}^{a_{ij}}(\theta) \ell_i(s_{i,t+1} | \theta)}{\sum_{\theta_k \in \Theta} \prod_{j=1}^n \mu_{j,t}^{a_{ij}}(\theta_k) \ell_i(s_{i,t+1} | \theta_k)}, \quad \theta \in \Theta. \quad (5)$$

The updating rule (5) is a distributed algorithm whereby each agent first aggregates the beliefs of its neighbors and itself as a new prior belief via weighted geometric average, and then uses Bayes' law to compute posterior distribution. For simplicity, in this paper we call the rule (5) **LoAB** (Logarithmic Aggregation and Bayesian update). The rule **LoAB** has been originally proposed and studied by Nedić *et al.* in [10], and the sufficient condition under which agents can learn the underlying true state using (5) is summarized as follows:

Condition 1:

- (1) *The time-varying network is B-strongly connected, i.e., there is an integer $B \geq 1$ such that the network is jointly strongly connected across every B time slots.*
- (2) *Any positive weight has a constant lower bound $\eta > 0$, i.e., if $a_{ij} > 0$ then $a_{ij} \geq \eta$ for all $i, j \in V$.*
- (3) *All agents have positive self-weights, i.e., $a_{ii} > 0$ for all $i \in V$.*
- (4) *All agents have positive initial belief on θ^* .*
- (5) *The true state θ^* is globally identifiable.*

In [10], Nedić *et al.* further consider a more general case where the true state θ^* might not be listed as one of the possible states. They define a set of state Ω_i for each agent i , where $\Omega_i = \arg \min_{\theta_k \in \Theta} D_{KL}(\ell_i(\cdot | \theta^*) \| \ell_i(\cdot | \theta_k))$, and let $\Omega^* \triangleq \cap_{i=1}^n \Omega_i$ which is not empty. That is to say, even though the true state might not be considered as a possible state, there exist some states which best explain the observations from the point of view of all agents. They prove that under such a relaxed condition, $\mu_{i,t}(\theta) \rightarrow 0$ almost surely as $t \rightarrow \infty$ for all i and $\theta \notin \Omega^*$. This implies that if there is only one state in Ω^* , all agents eventually assign belief of one on this state which is closest to the true state.

Next we propose an alternative way to introduce neighbors' information into the Bayesian update (2): instead of computing the weighted average of all KL-divergences between π and prior beliefs like that in (3), we can compute the KL-divergence between π and the weighted average of neighbors'

prior beliefs. The new optimization problem is as follows

$$\mu_{i,t+1} = \arg \min_{\pi \in \mathbb{P}(\Theta)} \left\{ D_{KL} \left(\pi \parallel \sum_{j=1}^n a_{ij} \mu_{j,t} \right) - \sum_{\theta_k \in \Theta} \pi(\theta_k) \log(\ell_i(s_{i,k+1} | \theta_k)) \right\}. \quad (6)$$

The solution of (6) is the Bayesian posterior distribution corresponding to the prior $\sum_{j=1}^n a_{ij} \mu_{j,t}$, i.e., for any $\theta \in \Theta$

$$\mu_{i,t+1}(\theta) = \frac{\sum_{j=1}^n a_{ij} \mu_{j,t}(\theta) \ell_i(s_{i,t+1} | \theta)}{\sum_{\theta_k \in \Theta} \left(\sum_{j=1}^n a_{ij} \mu_{j,t}(\theta_k) \ell_i(s_{i,t+1} | \theta_k) \right)}. \quad (7)$$

The essential difference between (5) and (7) is that the former aggregates prior beliefs via geometric average while the latter via arithmetics average. Here we call the algorithm (7) **LiAB** (Linear Aggregation and Bayesian update). We simply propose this algorithm in this paper without strict theoretical analysis. To the best of our knowledge, this algorithm has not been proposed and studied in the existing papers. Therefore, theoretically analyzing its performance is still an open question.

2.3 Aggregation of Bayesian Posterior Distributions

A common feature of the algorithms **LiAB** and **LoAB** is that they first aggregate local prior beliefs linearly or logarithmically as a new prior and then use the Bayes' law to compute the posterior distribution. We might change the order of the two steps and obtain two new algorithms.

If we let each agent first update its belief distribution based on its private observation and then exchange the posterior distribution with its neighbors via weighted geometric average, we obtain the following algorithm which is called **BLoA** (Bayesian update and Logarithmic Aggregation)

$$\mu_{i,t+1}(\theta) = \frac{\prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta)}{\sum_{\theta_k \in \Theta} \prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta_k)} \quad (8)$$

where

$$\tilde{\mu}_{i,t+1}(\theta) = \frac{\mu_{i,t}(\theta) \ell_i(s_{i,t+1} | \theta)}{\sum_{\theta_k \in \Theta} \mu_{i,t}(\theta_k) \ell_i(s_{i,t+1} | \theta_k)}. \quad (9)$$

The denominator in (8) is added to ensure that $\mu_{i,t+1}$ is still a well-defined probability distribution over Θ .

The rule (8) has been proposed and extensively studied in [11]. The sufficient condition for detecting the true state is summarized as follows:

Condition 2:

- (1) *The network is strongly connected.*
- (2) *All agents have positive initial belief on all $\theta \in \Theta$.*
- (3) *For every pair $\theta_p \neq \theta_q$, there is at least one agent i for which the KL-divergence $D_{KL}(\ell_i(\cdot | \theta_p) \| \ell_i(\cdot | \theta_q)) > 0$.*

Compared with *Condition 1*, the terms in *Condition 2* are more stringent. The second term requires that not only the initial beliefs of all agents on the true state are positive, but

also beliefs on all other states must be positive. The third term implies that there is no state that is observationally equivalent to any other state from the point of view of all agents in the network, i.e., all states are globally identifiable.

Note that the rule in [11] is not in the form like (8) but in the following form

$$\mu_{i,t+1}(\theta) = \frac{\exp\left(\sum_{j=1}^n a_{ij} \log \tilde{\mu}_{j,t+1}(\theta)\right)}{\sum_{\theta_k \in \Theta} \exp\left(\sum_{j=1}^n a_{ij} \log \tilde{\mu}_{j,t+1}(\theta_k)\right)}. \quad (10)$$

In fact, the rule (10) is identical to (8) since

$$\begin{aligned} & \frac{\exp\left(\sum_{j=1}^n a_{ij} \log \tilde{\mu}_{j,t+1}(\theta)\right)}{\sum_{\theta_k \in \Theta} \exp\left(\sum_{j=1}^n a_{ij} \log \tilde{\mu}_{j,t+1}(\theta_k)\right)} \\ &= \frac{\exp\left(\sum_{j=1}^n \log \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta)\right)}{\sum_{\theta_k \in \Theta} \exp\left(\sum_{j=1}^n \log \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta_k)\right)} \\ &= \frac{\exp\left(\log \prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta)\right)}{\sum_{\theta_k \in \Theta} \exp\left(\log \prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta_k)\right)} \\ &= \frac{\prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta)}{\sum_{\theta_k \in \Theta} \prod_{j=1}^n \tilde{\mu}_{j,t+1}^{a_{ij}}(\theta_k)}. \end{aligned}$$

From the above derivation it is not hard to understand why we call rules (5) and (8) logarithmic aggregations, since both of them involve geometric averages which can be written in logarithmic forms like that in (10).

Similar to (8) in the sense of aggregating Bayesian posterior beliefs of neighbors, an algorithm with geometric average being replaced by arithmetic average is proposed and extensively studied in [12], which is called **BLiA** (Bayesian update and Linear Aggregation) here

$$\mu_{i,t+1}(\theta) = \sum_{j=1}^n a_{ij} \tilde{\mu}_{j,t+1}(\theta) \quad (11)$$

where $\tilde{\mu}_{j,t+1}(\theta)$ is identical to that in (9). The sufficient conditions under which agents can detect the true state can be summarized as follows:

Condition 3:

- (1) The weight matrix $A = [a_{ij}]$ is primitive.
- (2) There exists at least one agent with positive initial belief on the true state.
- (3) For each agent i , there exists at least one prevailing signal s_i^o such that $\ell_i(s_i^o|\theta^*) - \ell_i(s_i^o|\theta) \geq \delta_k^o > 0$ for any $\theta \notin \Theta_i$.

Compared with *LoAB* and *BLoA*, the algorithm *BLiA* requires more relaxed conditions in some aspects to detect the underlying true state. For instance, it only needs at least one agent having positive initial belief on the true state. And also, the requirement of primitive matrix is more relaxed than that with positive diagonal elements (i.e., positive self-weights).

2.4 Aggregation of Personal Bayesian Posterior and Others' Prior

The following two algorithms, like *BLoA* and *BLiA*, also contain personal Bayesian update and communication with neighbors. However, agents exchange prior distributions with their neighbors rather than posterior distributions. We may consider that this sort of algorithms introduce communication delay into the dynamics such that agents could not receive their neighbors' latest information (the posteriors) but only delayed information (the priors before Bayesian update).

There are also two ways to aggregate neighbors' information. If the agent aggregate its personal posterior and its neighbors' priors via arithmetic average, we have the following algorithm called **BLiAD** (Bayesian update and Linear Aggregation of Delayed information)

$$\mu_{i,t+1}(\theta) = a_{ii} \tilde{\mu}_{i,t+1}(\theta) + \sum_{j \in N_i} a_{ij} \mu_{j,t}(\theta). \quad (12)$$

The rule (12) is originally proposed in [13] in the context of social learning. The sufficient condition under which agents can learn the underlying true state using (12) can be summarized as follows:

Condition 4:

- (1) The network is strongly connected.
- (2) All agents have strictly positive self-weights, i.e., $a_{ii} > 0$ for all $i \in V$.
- (3) There exists at least one agent with positive initial belief on the true state θ^* .
- (4) The true state θ^* is globally identifiable.

Replacing the arithmetic average in *BLiAD* by the geometric average, we obtain another algorithm here called **BLoAD** (Bayesian update and Logarithmic Aggregation of Delayed information) as follows

$$\mu_{i,t+1}(\theta) = \frac{\tilde{\mu}_{i,t+1}^{a_{ii}}(\theta) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta)}{\sum_{\theta_k \in \Theta} \left(\tilde{\mu}_{i,t+1}^{a_{ii}}(\theta_k) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta_k) \right)}. \quad (13)$$

Similar to (13), in [14] Rad and Tahbaz-Salehi propose a distributed estimation algorithm as follows:

$$\log \mu_{i,t+1}(\theta) = \lambda_i \log \ell_i(s_{i,t+1}|\theta) + \sum_{j=1}^n a_{ij} \log \mu_{j,t}(\theta) + c_{i,t} \quad (14)$$

where $\lambda_i > 0$ is the weight that agent i assigns to its private observations and $c_{i,t}$ is a normalization constant, not dependent on θ , which ensures that $\mu_{i,t+1}$ is a well-defined probability distribution over Θ .

The rule (14) is identical to (13) by choosing the following values for the parameters λ_i and $c_{i,t}$:

$$\lambda_i = a_{ii}$$

and

$$c_{i,t} = -\log \left[\left(\sum_{\theta_k \in \Theta} \mu_{i,t}(\theta_k) \ell_i(s_{i,t+1}|\theta_k) \right)^{a_{ii}} \times \sum_{\theta_k \in \Theta} \left(\tilde{\mu}_{i,t+1}^{a_{ii}}(\theta_k) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta_k) \right) \right].$$

In fact, the rule (14) can be written in the following form:

$$\begin{aligned} & \mu_{i,t+1}(\theta) \\ &= \exp \left(\lambda_i \log \ell_i(s_{i,t+1}|\theta) + \sum_{j=1}^n a_{ij} \log \mu_{j,t}(\theta) + c_{i,t} \right) \\ &= \exp \left(\log \ell_i^{\lambda_i}(s_{i,t+1}|\theta) + \log \prod_{j=1}^n \mu_{j,t}^{a_{ij}}(\theta) + c_{i,t} \right) \\ &= \ell_i^{\lambda_i}(s_{i,t+1}|\theta) \prod_{j=1}^n \mu_{j,t}^{a_{ij}}(\theta) \cdot e^{c_{i,t}} \\ &= \ell_i^{a_{ii}}(s_{i,t+1}|\theta) \mu_{i,t}^{a_{ii}}(\theta) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta) \cdot e^{c_{i,t}} \\ &= \frac{\tilde{\mu}_{i,t+1}^{a_{ii}}(\theta) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta)}{\sum_{\theta_k \in \Theta} \left(\tilde{\mu}_{i,t+1}^{a_{ii}}(\theta_k) \prod_{j \in N_i} \mu_{j,t}^{a_{ij}}(\theta_k) \right)}. \end{aligned} \quad (15)$$

The rule (14) has been theoretically studied in [14], and the sufficient condition for detecting the true state is summarized as follows:

Condition 5:

- (1) The network is strongly connected.
- (2) All agents have positive initial prior belief on all $\theta \in \Theta$.
- (3) The true state θ^* is globally identifiable.

As a summarization, we provide the classification tree of the six distributed detection algorithms in Fig. 1.

2.5 Comparison with the Bayesian Update of a Single Agent

L. J. Savage has pointed out in [17] that, barring two banal exceptions, a single agent becomes almost certain of the true state when the amount of its observation increases infinitely. One exception is that the initial belief of the true state is zero. This is very easy to understand. If the belief is zero, then, no matter what signal is observed, the posterior belief of the true state is still zero. The other exception occurs when there exists a state which arouses exactly the same signals as the true state does, i.e., observationally equivalent state exists.

When an agent is situated in a network setting, the above two requirements might be relaxed. For instance, it has been proven that the rules *LoAB*, *BLoA*, *BLiA*, *BLiAD*, and *BLoAD* only require the true state being globally identifiable. We conjecture that *LiAB* might also work well with the same condition, even though theoretical analyses are not available yet.

It is not hard to see that all rules involving geometric averages, such as *LoAB*, *BLoA*, and *BLoAD*, still need the requirement of non-zero initial beliefs of all agents on the true state.

For the rules *BLiA* and *BLiAD* which contains linear aggregations, it has been proven in [12] and [13], respectively, that at least one agent with non-zero initial belief on the true state is enough for a correct detection.

For the rules involving delayed information such as *BLiAD* and *BLoAD*, the requirement of non-zeros self-weights must be satisfied, at least for part of the agents. Because if all self-weights are zeros, any new observation will be discarded and *BLiAD* and *BLoAD* specialize to traditional consensus protocols with arithmetic average and geometric average, respectively.

3 Numerical Examples

In this section, we exam the effectiveness of each distributed detection rule by numerical examples. Our test platform is a nearest-neighbor coupled network, which might represent a sensor network or a robot network where, restricted by its communication range, each agent can only exchange information with a given number of its closest neighbors. The following Fig. 2 shows a schematic diagram of a nearest-neighbors coupled network of 20 agents, where each agent could only interact with 5 closest agents (including itself) and all of the edges are bi-directed.

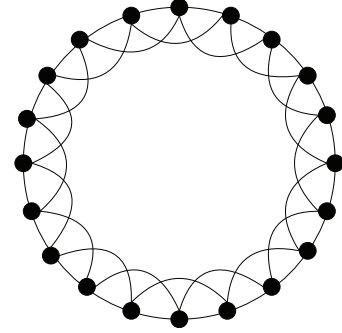


Fig. 2: A nearest-neighbor coupled network of 20 agents

Simulations are performed on three possible states, i.e., $\Theta = \{\theta_1, \theta_2, \theta_3\}$ in which θ_3 is set to be the true state. Any agent i 's belief distribution at time t is $\mu_{i,t} = [\mu_{i,t}(\theta_1), \mu_{i,t}(\theta_2), \mu_{i,t}(\theta_3)] \in [0, 1]^3$. The initial belief is uniformly distributed in the interval $[0, 1]$ and subject to $\sum_{k=1}^3 \mu_{i,t}(\theta_k) = 1$.

The signals generated by the true state are $\{s_1, s_2\}$. We assume that signal s_1 appears with possibility of 0.8, and s_2 with 0.2, which implies the private signal structure about θ_3 of any agent i should be $\ell_i(s_1|\theta_3) = 0.8$ and $\ell_i(s_2|\theta_3) = 0.2$.

Let half of agents, denoted by $V_1 \subset V$, have the signal structures $\ell_i(s_1|\theta_1) = 0.8$, $\ell_i(s_2|\theta_1) = 0.2$, $\ell_i(s_1|\theta_2) = 0.5$, and $\ell_i(s_2|\theta_2) = 0.5$ ($i \in V_1$). That is to say, the states θ_1 and θ_3 are equivalent to agents in V_1 . Let the other half of agents, denoted by $V_2 = V \setminus V_1$, have the following signal structures: $\ell_i(s_1|\theta_1) = 0.2$, $\ell_i(s_2|\theta_1) = 0.8$, $\ell_i(s_1|\theta_2) = 0.8$, and $\ell_i(s_2|\theta_2) = 0.2$ ($i \in V_2$), i.e., the state θ_2 and θ_3 are observationally equivalent to agents in V_2 . The true state is unidentifiable to any single agent, but globally identifiable.

In our first simulation, we let agents from V_1 be close to each other, and the same to V_2 , i.e., agents belong to the same set form a cluster. In the second simulation, we mix all agents in the sense that each agent from V_1 is located between two agents from V_2 , and vice versa.

We focus on the number of iterations of update for each algorithm to detect the true state. If for all $i \in V$, $|\mu_{i,t}(\theta_3) - 1| \leq 10^{-3}$, we say the whole group collectively detect the true state. The result is shown in Fig. 3, which is the average of 100 realizations. From Fig. 3, we have the following observations:

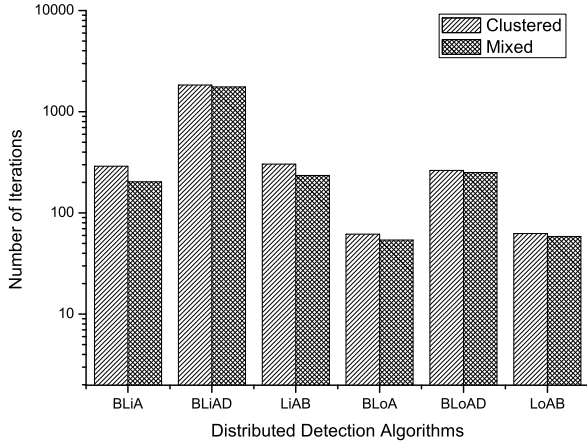


Fig. 3: Comparison of six distributed detection algorithms

- (1) All of the six distributed detection algorithms are effective in detecting the true state.
- (2) If other aspects are identical, aggregating information via geometric average (i.e., $BLoA$, $BLoAD$, and $LoAB$) is much faster than that via arithmetic average (i.e., $BLiA$, $BLiAD$, and $LiAB$).
- (3) Using delayed information in aggregation (i.e., $BLiAD$ and $BLoAD$) makes the detection much slower.
- (4) Bayesian update first or aggregating information from neighbors first does not influence the efficiency of detection apparently.
- (5) Mixing agents with different signal structures promotes the rate of detection.

The above items (2) and (3) are in accordance with the theoretical result in [11] where $BLoA$ is compared with $BLiAD$. One of the main results is that, the lower bound on the rate of detection by using $BLoA$ is even greater than the upper bound of $BLiAD$. The item (3) is also in accordance with the theoretical result in [12], in which $BLiA$ is compared with $BLiAD$, and the former is much faster in detecting the true state.

4 Concluding Remarks

In this paper, we discuss a class of distributed detection algorithms in which personal Bayesian updates are combined with some types of consensus protocols. We focus on six algorithms and classify them according to the type of consensus protocol,

the order of Bayesian update and consensus, and whether time-delayed information is involved in the interaction. By comparison, we have a systematic impression of these distributed detection algorithms, which might lead us to establishing more refined conditions under which agents could detect the underlying true state in distributed settings. For instance, could the terms (2) and (3) in *Condition 2* be replaced by more relaxed requirements, say the terms (4) and (5) in *Condition 1*, respectively? And also, could the *Conditions 2, 3, 4, and 5* be relaxed to time-varying networks like that in *Condition 1*? Through numeric examples, we obtain some qualitative results about the efficiency of these distributed algorithms, which might shed light on designing more efficient distributed detection algorithms in the future.

There are many other types of distributed detection algorithms proposed in the literature. For instance, instead of exchanging beliefs, agents can share their signal structures with their neighbors which could also result in a correct detection of the true state [7–9]. Also, Bayesian update is not the only choice in detection problem. More alternatives can be found in the literature (e.g., [18–21]).

References

- [1] M. O. Jackson, *Social and Economic Networks*. New Jersey: Princeton University Press, 2010.
- [2] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*. London: Springer-Verlag, 2008.
- [3] R. W. Beard, T. W. McLain, D. Nelson, D. Kingston and D. Johanson, Decentralized cooperative aerial surveillance using fixed-wing miniature UAVs, *Proceedings of the IEEE*, 94(7): 1306–1324, 2006.
- [4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, Randomized gossip algorithms, *IEEE Trans. on Information Theory*, 52(6): 2508–2530, 2006.
- [5] L. Smith and P. Sorensen, Pathological outcomes of observational learning, *Econometrica*, 68(2): 371–398, 2000.
- [6] D. Acemoglu and A. Ozdaglar, Opinion dynamics and learning in social networks, *Dynamics Games and Applications*, 1(1): 3–49, 2010.
- [7] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, Distributed detection : finite-time analysis and impact of network topology, *arXiv preprint arXiv:1409.8606v1*, 2014.
- [8] B. S. Rao and H. Durrant-Whyte, A decentralized Bayesian algorithm for identification of tracked targets, *IEEE Trans. on Systems, Man, and Cybernetics*, 23(6): 1683–1698, 1993.
- [9] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, Belief consensus and distributed hypothesis testing in sensor networks, in *Workshop on Network Embedded Sensing and Control*, 2006: 169–182.
- [10] A. Nedić, A. Olshevsky, and C. A. Uribe, Nonasymptotic convergence rates for cooperative learning over time-varying directed graphs, *arXiv preprint arXiv:1410.1977v1*, 2014.
- [11] A. Lalitha, T. Javidi, and A. Sarwate, Social learning and distributed hypothesis testing, *arXiv preprint arXiv:1410.4307v2*, 2014.
- [12] X. Zhao and A. H. Sayed, Learning over social networks via diffusion adaptation, in *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, 2012: 709–713.
- [13] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi,

- Non-Bayesian social learning, *Games and Economic Behavior*, 76(1): 210–225, 2012.
- [14] K. R. Rad and A. Tahbaz-Salehi, Distributed parameter estimation in networks, in *Proceedings of 49th IEEE Conference on Decision and Control*, 2010: 5050–5055.
- [15] A. Zellner, Optimal information processing and Bayes’ theorem, *The American Statistician*, 42(4): 278–280, 1988.
- [16] S. G. Walker, Bayesian inference via a minimization rule, *Sankhyā: The Indian Journal of Statistics*, 68(4): 542–553, 2006.
- [17] L. J. Savage, *The Foundations of Statistics*, New York: Wiley, 1954, chapter 6.
- [18] R. Viswanathan and P. K. Varshney, Distributed detection with multiple sensors: part I—fundamentals *Proceedings of the IEEE*, 85(1): 54–63, 1997.
- [19] R. S. Blum, S. A. Kassam, and H. V. Poor, Distributed detection with multiple sensors: part I—advanced topics *Proceedings of the IEEE*, 85(1): 64–79, 1997.
- [20] J. B. Predd, S. R. Kulkarni, and H. V. Poor, Distributed learning in wireless sensor networks, *IEEE Signal Processing Magazine*, 23(4): 56–69, 2006.
- [21] A. H. Sayed, *Adaptation, Learning, and Optimization over Networks*, Boston-Delft: NOW Publishers, 2014.

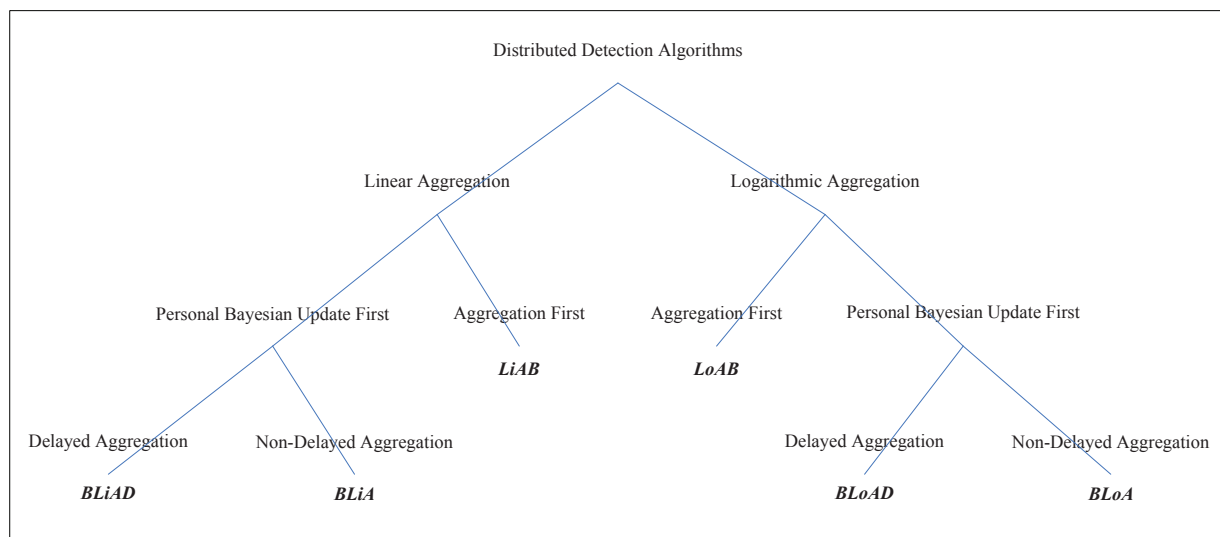


Fig. 1: Classification tree of the six distributed detection algorithms